Journal of Solid Waste Valorization and Chemical Management, 2026, Vol. 01

ARTICLE DOI: 10.66729/jswvcm202603002

A Dual-Strategy Deep Learning Framework for Automated Defect Detection: Enabling Efficient Echelon Utilization of Spent Lithium-Ion Batteries

Yuhui Peng[1], Yaxin Chi[1,2], Xihua Zhang[1*], Tao Zhang[2], Xuning Zhuang[1], Xiaolong Song[1]

1 School of Resources and Environmental Engineering, Shanghai Polytechnic University; Shanghai Collaborative Innovation Center for WEEE Recycling, Shanghai 201209, China.

2 North Star Advanced Recycling Technology (Qingdao) Co., Ltd., Shandong Qingdao 266109, China.

*Corresponding Author: Xihua Zhang; E-mail: zhangxh@sspu.edu.cn

Abstract

Echelon utilization, as the preferred method for reusing spent power batteries, relies on accurate detection of appearance defects to ensure safe and reliable operation. However, traditional methods rely heavily on manual detection, making it difficult to meet the demands of industrialization. This work proposes a dual mode deep learning framework for automated defect detection in spent power batteries, aiming to satisfy both real time inspection and high precision detection requirements. In this framework, a YOLOv5 model enhanced by transfer learning is employed for rapid real time detection, while a Mask R-CNN model improved with SoftNMS is developed for high precision detection in overlapping target scenarios. Compared with conventional NMS, SoftNMS reduces missed detections by continuously decaying the confidence scores of overlapping candidate boxes rather than directly suppressing them. To support model training and performance evaluation, a specialized dataset of appearance defects in spent batteries is constructed. Experimental results show that the improved YOLOv5 achieves a precision of 96.70%, while the improved Mask R-CNN reaches a precision of 97.44%. In addition, a PyQt5 based visual inspection system is developed to verify the practical applicability of the proposed framework in automated battery sorting. The results demonstrate that the proposed method provides an effective technical solution for intelligent defect detection in the echelon utilization of spent power batteries.

Keywords

Echelon utilization; Deep learning; Defect detection; YOLOv5; Mask R-CNN

Citation: Yuhui Peng, Yaxin Chi, Xihua Zhang and *et al.* A Dual-Strategy Deep Learning Framework for Automated Defect Detection: Enabling Efficient Echelon Utilization of Spent Lithium-Ion Batteries. *Journal of Solid Waste Valorization and Chemical Management*. 2026, 01 (1): 1-9. doi: 10.66729/jswvcm202603002.

1. Introduction

With the rapid expansion of new energy vehicles, the number of spent power batteries has increased, and given their inherent 5-10 year lifespan, the need for effective recycling and utilization has become increasingly urgent. At present, echelon utilization represents the most ecologically friendly and energy-efficient method for recycling and reusing spent power batteries. It is also the preferred utilization method to reduce the environmental impact and carbon emissions for the power battery industry[1,2]. During this process, appearance detection is a crucial step to ensure safety and reliability of echelon utilization batteries. In 2021, China issued the national standard Automotive Power Battery Recycling and Echelon Utilization, including Part 3, Echelon Utilization Requirements, which specifies the appearance requirements for batteries intended for echelon utilization. In order to ensure the safety during echelon utilization, it is crucial to carry out the appearance tests for the batteries prior to echelon utilization. At present, defect detection relies on manual detection, which fails to meet the demands for fast and effective data processing and is prone to missed detection, false detection, and other issues[3]. Manual detection also lacks scalability for industrial production lines. Therefore, it is urgent to develop intelligent detection technology and automated detection equipment.

In recent years, deep learning based appearance defect detection has been extensively studied in various industrial contexts. These applications include steel appearance defect detection[4], metal casting defect detection[5], appearance defects in ceramics[6,7], tunnel defect detection[8]. These studies have demonstrated the effectiveness of deep learning models such as YOLO and Mask R-CNN in achieving high detection precision and real-time performance. However, most of the existing studies are based on two-dimensional (2D) planar images. For example, Chen applied an improved YOLOv5 model to solar cell defect detection[9], and Wang investigated defect detection on pipeline defects[10]. These methods are difficult to apply directly to three-dimensional (3D) images with curved appearance structures. The appearance texture of cylindrical batteries are complex and change with the curved appearance, which limits the 2D feature extraction method. In addition, appearance defects of batteries often present the characteristics of small size and unclear edges, and are more prone to deformation on 3D curved appearances, making the identification of tiny detects a current technical difficulty. Compounding this issue, public datasets for the type of 3D appearance defects are also relatively scarce. Therefore, how to achieve high-precision detection of minute defects in 3D cylindrical battery images remains a key

Copyright:© 2026 The Author(s). Published by ICJN Press.
This is an open access article under the https://creativecommons.org/licenses/by/4.0/
Received 13 March 2026; Revised 21 April 2026; Accepted 14 May 2026; Available online 28 May 2026.

challenge in current research.

These practical challenges arise from a fundamental mismatch between current models and the way defects are manifested on 3D curved appearances. Traditional image-processing methods provided limited flexibility because their manually designed features could not adapt to diverse defect characteristics. Although deep learning has improved feature representation, its learning mechanisms are still largely tailored to 2D planar imagery. When applied to cylindrical batteries, the feature extraction process no longer aligns naturally with the way defects manifest on these geometries. This misalignment between defect characteristics and model learning mechanisms makes the recognition of minute scratches on 3D appearances particularly challenging. Therefore, solving the adaptation problem between deep learning models and 3D defect characteristics are crucial for improving the precision of tiny detection on cylindrical battery appearance.

To address the major challenges in appearance defect detection for spent cylindrical power batteries, this work develops an integrated framework with explicit correspondence between detection problems and technical solutions. For the problems of curved appearance and reflective interference, a dedicated image acquisition environment is established according to the reflective characteristics of battery steel shells, and a high resolution dataset of 18,650 and 21,700 batteries is constructed to provide high quality image data for model training and evaluation. For industrial scenarios requiring rapid online inspection, a YOLOv5 model enhanced by transfer learning is adopted to achieve efficient real time detection. For tiny defects and overlapping targets that are prone to missed detection, a Mask R-CNN model optimized with the SoftNMS algorithm is introduced to improve high precision recognition performance. In addition, a PyQt5 (version not recorded; Riverbank Computing Limited, Dorchester, England, UK, with Qt 5.15.2 as the underlying framework) based visualization and monitoring system is developed to support practical deployment and real time interaction. The proposed framework therefore provides a targeted solution for intelligent defect inspection in spent power batteries and also shows potential for production-side cell screening and quality categorization.

2. Methods

2.1 Workflow and Strategy

To address the dual challenges of real time detection and high precision detection in battery echelon utilization, this work designs a hybrid detection framework (**Figure. S1**). The appearance defect detection task for spent cylindrical batteries has several distinctive characteristics. First, the steel shell surface is curved and reflective, which increases image variation and makes defect features more sensitive to illumination and viewing conditions. Second, typical defects such as scratches, stains, pits, and leakage traces are often tiny, irregular, and locally distributed, making them difficult to detect stably. Third, practical sorting scenarios require not only fast online inspection, but also fine identification of subtle defects for high standard quality control. Deep learning algorithms offer higher precision and stronger robustness[11,12]. But different model architectures are suited to different inspection objectives. Single-stage defect detection methods are particularly suitable for scenarios requiring fast and real-time defect detection[13,14]. However, existing studies have shown that such models may suffer from insufficient recall and sensitivity to imaging conditions. Although lightweight or transfer learning based improvements have been explored, including compressed YOLOv5 architectures and enhanced feature extraction backbones[15,16], their performance is still unstable when training data are limited and the target surface is curved and reflective. Therefore, this work selects YOLOv5 as the real time detection model and further introduces transfer learning, so that the model can better adapt to the domain specific appearance characteristics of spent cylindrical batteries and achieve more robust screening performance under complex image acquisition conditions. For high precision inspection, appearance defect detection methods based on two stage detection provide stronger feature representation and localization capability[17-20]. Previous studies, such as Cha and Xie, have demonstrated the effectiveness of Faster R-CNN and Fast R-CNN in structural defect detection[21,22]. Compared with single stage models, the two stage mechanism is more suitable for the identification of tiny, subtle, and irregular defects, which is particularly important for curved battery appearances where local defect features are easily weakened. On this basis, Mask R-CNN is selected in this work because it not only performs defect detection, but also provides pixel level defect region description through its mask branch, which is beneficial for refined defect characterization. Furthermore, to reduce missed detection caused by overlapping candidate regions, the traditional NMS algorithm is replaced with SoftNMS. This strategy preserves valid candidate boxes by decaying their confidence scores rather than directly removing them, thereby improving the recognition of densely distributed or partially overlapping defects without increasing model complexity.

2.2. Data Collection and Preprocessing

This work selected 18,650 and 21,700 cylindrical batteries, which were the most used in new energy vehicles, as the primary work objects. The five common defects of breakage, pit, stain, leakage, and scratch were taken as the primary defect objects in this work, as shown in the **Figure. 1 (a)**.

Deep learning relies on a large amount of data as input data and supervised learning to absorb and summarize prior knowledge[23]. Therefore, high-quality image acquisition is a critical step in appearance defect detection. However, the external cladding and internal steel shell of batteries are highly reflective. This challenge is particularly pronounced for cylindrical batteries, where curved metallic appearances tend to cause non-uniform reflections and local highlights during image acquisition. To address this issue, only side lights are used for illumination in this work. As shown in the **Figure. 1 (b)**, translucent polypropylene (PP) panels are placed on the left, right, and front sides of the battery to be collected to form a light shield. By employing uniformly scattered illumination, strong reflections from the battery appearance are effectively suppressed, thereby enabling the acquisition of high-quality image data.

The battery samples used in this work were supplied by a cooperating battery recycling enterprise in Suzhou and were used for image acquisition after simple pretreatment. Due to the limited availability of qualified spent battery samples and the screening requirements for defect identification, a total of 625 original images were finally retained. Each image contained at least one type of defect, and all images were stored at a uniform resolution of 3,472 × 3,472 pixels. Some representative images from the dataset are shown in the **Figure. S2**. Image acquisition is the first step in constructing the appearance defect dataset. In addition to image data, deep learning also depends on high quality labeled annotations to provide the prior knowledge required for supervised learning. Among these factors, data annotation plays a critical role, as both the quantity and quality of training samples substantially influence model performance[24]. As shown in the **Figure. 1 (c)**, the polygonal tracing method in LabelMe[25] was used to annotate the batteries and their appearance defects. To

compensate for the limited scale of the retained original images, data augmentation was applied to the original images, and a detailed description of the augmentation strategy is provided in **Supplementary Information (SI. 3).**

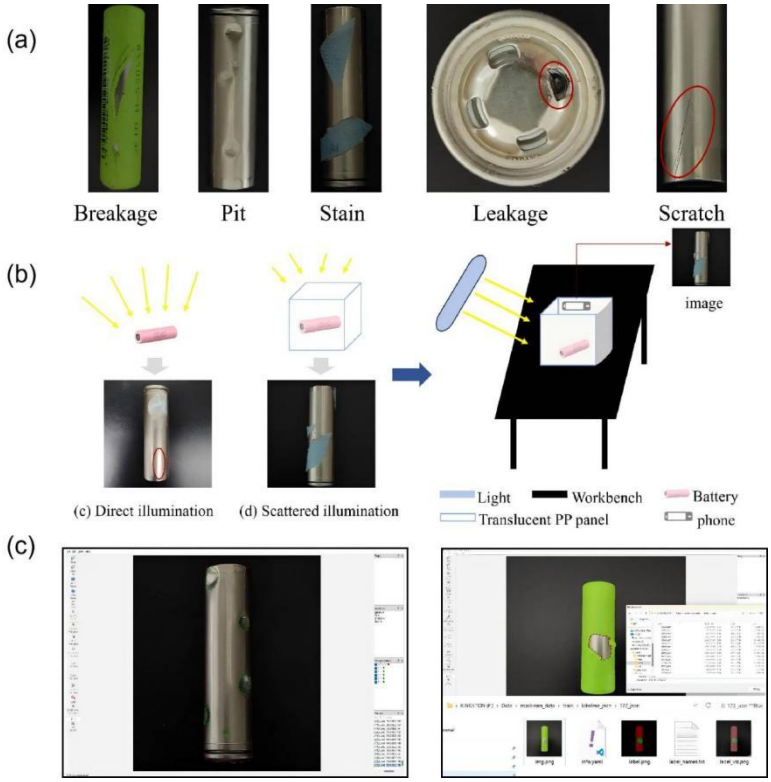

Figure. 1 Defect categories and dataset construction process for cylindrical batteries. **(a)** Types of battery appearance defects; **(b)** The method of light source projection; **(c)** Dataset image label annotation interface. Produced by Powerpoint (version 12.1.0.25865).

After data augmentation, the image dataset used for model training was constructed according to the requirements of the algorithms employed in this work. The dataset structure corresponding to the two models is shown in the **Figure. S4**. A total of 3,750 images were divided into training, validation, and test sets at a ratio of 8:1:1, resulting in 3,000 training images, 375 validation images, and 375 test images. To avoid direct information leakage, each original image and all of its augmented versions were assigned to the same dataset partition. Since the battery samples were provided by the cooperating enterprise, the statistics of the data set in this work are based on the number of retained images.

2.3. Machine Learning

YOLOv5 is a single-stage target detection network that can simultaneously classify and localize targets by extracting features directly from images. Its high detection speed enables real-time performance, making it particularly suitable for scenarios that demand rapid response. In this work, the YOLOv5s-seg model was selected as a representative single-stage network for experimental analysis. The structural diagram and detailed description of the YOLOv5 architecture are provided in the **Supplementary Material (SI. 5)**. Because the YOLOv5 model showed unsatisfactory performance in detecting appearance defects, particularly in identifying scratch defects, transfer learning was introduced to improve its detection performance. Transfer learning is a significant machine learning technique that leverages knowledge acquired from one task to improve learning efficiency and performance on related tasks[26]. In transfer learning, the source domain is the domain of existing knowledge, and the target domain is the domain that needs to be learned, which is the dataset constructed in this work for cylindrical battery defect detection. In this work, the appearance defect dataset (NEU-DET dataset) of Northeastern University was used as the source domain dataset for the experiments[19,27,28]. This dataset contains 1,800 images, covering six typical appearance defect categories of hot-rolled strip steel. Although the detection objects in the NEU-DET dataset are different from those in the target domain of this work, both of these tasks belong to industrial appearance defect detection and have similar underlying visual features. Specifically, the defects in these two areas exhibit similar texture variations, discontinuous edges and irregular contours. These features correspond to the low-level and intermediate features learned by the early convolutional layers of YOLOv5. Based on this observation, the YOLOv5s model was first pre-trained on the NEU-DET dataset to obtain defect-sensitive feature representations. The pre-trained weights, learned from the source domain data, are transferred to the target domain model to initialize the network parameters. This transfer of learned weights allows the model to leverage prior knowledge of texture and edge features, which helps improve detection performance for scratch defects in cylindrical battery images, despite the small size and limited diversity of the target domain dataset. Then, the optimal parameters and weights obtained from its training are used for the initialization of the target domain model, and then the model parameters are fine-tuned using the target domain dataset. Through this transfer learning strategy, the model benefits from the prior knowledge learned from industrial appearance defect data, thereby improving the detection performance under the condition of limited target domain data. Based on this, the training process was organized into three stages: pre-training, transfer learning, and validation, as shown in the **Figure. S6**. As a single-stage detection model, YOLOv5 offers exceptionally fast detection speed. However, its performance in identifying tiny targets remains

unsatisfactory. The Mask R-CNN model is employed to address this, it is better suited for scenarios requiring high detection precision. Mask R-CNN is a deep learning model for object detection and instance segmentation. By adding branches, the model can capture the precise boundaries of defect targets through pixel-level masks, enabling more accurate defect identification. As a result, this work adopted the Mask R-CNN framework for appearance defect detection. Mask R-CNN is a composite detection architecture that integrates region proposal, object classification, bounding-box regression, and instance segmentation into a unified multi-branch network. The structural diagram and detailed description of the Mask R-CNN model were provided in **Supplementary Material (SI. 7)**. This work focuses on spent power batteries, which often exhibit complex appearance defects and overlapping defects. Under such conditions, the Mask R-CNN model may miss tiny defects, such as scratches, due to interference in the background and suppression of overlapping candidates. To mitigate this issue, the model was modified by replacing the NMS algorithm with the SoftNMS algorithm, enabling the model to retain plausible candidates rather than discarding them abruptly. This adjustment improved the model's ability to detect tiny defects in cluttered environments. Further details and algorithmic formulations are provided in the **Supplementary Information (SI. 8)**.

3. Results and Discussion

Detailed implementation settings, training hyperparameters, and evaluation indicators of YOLOv5s and Mask R-CNN are provided in **Supplementary Information (SI. 9)** including the hardware environment, software configuration, learning rate, batch size, epoch, weight decay, momentum, NMS threshold, and warm-up settings. Unless otherwise specified, the remaining untuned parameters follow the default settings of the adopted framework.

3.1. Rapid Real-time Detection Model Developed and Adaptability

Based on the experimental setup, the labeled dataset was put into the YOLOv5s network for training. A total of 3,000 images were used in the training set, with the model initialized using the default pre-trained weights, and the visualization tool TensorBoard (version 2.10.1; developed by the Google Brain Team, Google LLC, Mountain View, California, USA) was used to track the evaluation indicators of the model.

After 300 epochs of training, the model detected the batteries and their appearance defects in all images of the training set and generated corresponding category labels and rectangular bounding boxes for each target. **Figure. 2 (a)** statistically presents the distribution of different defect categories in the training set. The results show an obvious class imbalance, with scratch defects accounting for the largest proportion, while leakage defects are relatively few. This distribution reflects the practical defect characteristics of spent cylindrical batteries and indicates that the model must address class imbalance during training. **Figure. 2 (c)** shows the spatial distribution of defect annotations, where the x and y axes are normalized coordinates relative to the image size, ranging from 0 to 1. The results indicate that most defects are concentrated in the central region of the images. **Figure. 2 (d)** presents the distribution of defect sizes relative to the image size, where both Width and Height are normalized values ranging from 0 to 1. It can be seen that most defects belong to tiny targets. These statistical results indicate that the dataset is characterized by class imbalance, a high proportion of tiny targets, and a relatively concentrated spatial distribution.

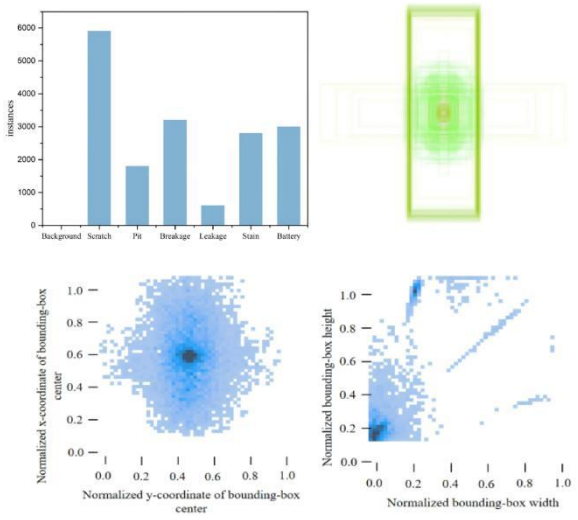

Figure. 2 Statistical plot of a training set data. **(a)** Statistical results of the number of batteries and their different types of appearance defects in the training set images; **(b)** Statistics of all batteries and their appearance defects labeled boxes in the training set images; **(c)** The central point of all defects in each picture corresponds to the position of the whole picture; **(d)** Statistics of defect size as a percentage of the entire image. Produced by Origin (version 9.8.0.200).

Figure. 3 presents the average precision (AP) values for each defect category. The model achieves high precision in detecting Battery, Stain, Leakage, and Pit defects. However, the detection performance for scratches is markedly lower, with an AP of only 17.7%. This result indicates that the YOLOv5s model trained with the initial weights is unable to learn sufficiently discriminative features for scratch defects, leading to frequent missed detection and instability in localization. Scratch defects on cylindrical battery appearances are typically characterized by extremely tiny sizes, elongated shapes, and weak texture contrast with the background. During training from scratch, the feature extraction layers of YOLOv5s tend to prioritize salient patterns with stronger visual responses, such as pits or stains, whose features are more easily captured by convolutional filters. In this situation, the subtle visual cues associated with scratches are not captured well. As a consequence, although the model demonstrates satisfactory performance on defect categories with clearer geometric and textural characteristics, its ability to detect scratch defects remains limited. This observation highlights the necessity of enhancing the model's sensitivity to fine-scale features and providing more informative feature initialization to improve scratch detection performance.

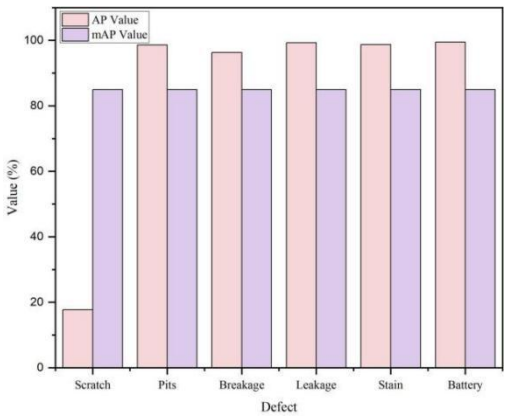

Figure. 3 AP values for cylindrical batteries and their various types of defects. Produced by Origin (version 9.8.0.200).

To enhance model performance, transfer learning was introduced. No dedicated class balancing strategy was used during training beyond the conventional data augmentation described in 2.2. Therefore, the class wise results mainly reflect the influence of sample distribution and defect characteristics. As shown in the **Figure. 4 (a)** and the **Figure. 4 (b)**, the AP for scratch defects increases markedly from 17.70% to 64.80% after transfer learning, while the AP for breakage defects rises from 96.30% to 98.90%, and the other major categories show only minor changes. This indicates that transfer learning is particularly beneficial for scratch defects, which have relatively limited samples and subtle visual features. In contrast, the leakage category shows only a slight change because its baseline AP is already high. **Figure. 4 (c)** shows the loss convergence before and after optimization. The baseline YOLOv5s first reaches a loss of 0.005 at the 47th epoch and converges between 0.003 and 0.004, whereas the transfer YOLOv5s reaches 0.005 at the 26th epoch and converges between 0.001 and 0.002. The faster decline and smoother convergence indicate better feature reuse and optimization stability after transfer learning.

As shown in the **Figure. 4 (d)**, transfer learning also improves the overall test performance. The precision increases from 95.40% to 96.70%, the recall rises from 82% to 84.30%, and the mAP improves from 85% to 93.40%. Meanwhile, the detection speed decreases only slightly from 200 to 179 FPS. These results show that transfer learning improves detection accuracy while maintaining high efficiency.

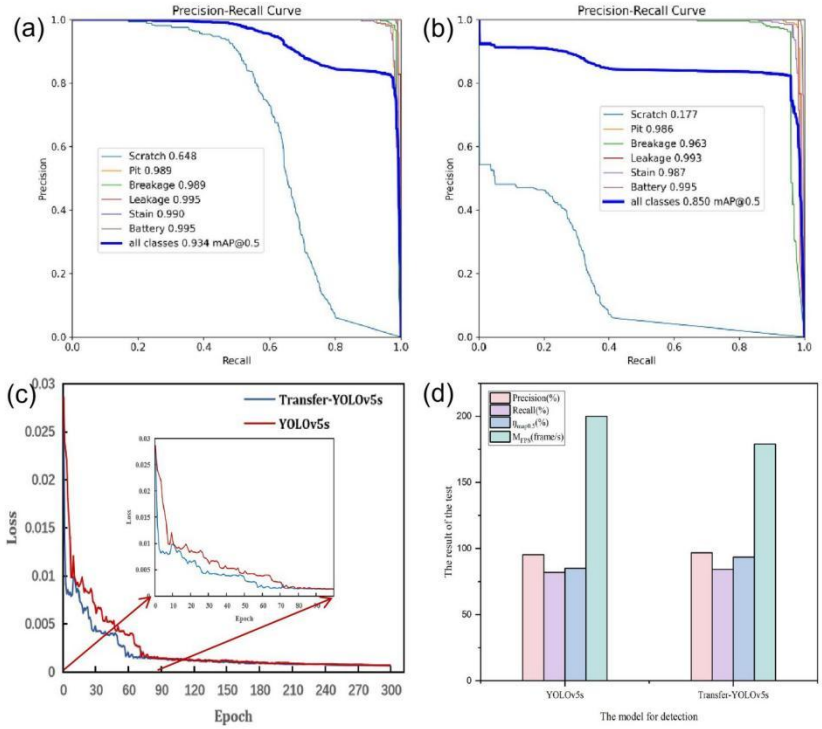

Figure.4 Comparison of the models before and after optimization. **(a)** The P-R curve of the model before optimization; **(b)** The P-R curve of the model after optimization; **(c)** Comparison chart of the changes in loss functions before and after optimization of the two models; **(d)** Network detection results with different weight initialization. Produced by Python (version 3.7) and Powerpoint (version 12.1.0.25865).

The detection results on the test set are shown in the **Figure. 5**. As observed, the optimized model achieves more accurate localization and higher confidence for representative defect categories. For stain, the confidence score increases from 0.49 to 0.66, with fewer missed regions along the edges. For the pit, the confidence rises from 0.39 to 0.83, indicating improved recognition of small circular defects in reflective areas. For scratch, the optimized model more precisely captures the narrow elongated region, with the confidence improving from 0.73 to 0.96. These qualitative results are consistent with the quantitative analysis and further confirm that transfer learning is especially beneficial for subtle defects with limited samples.

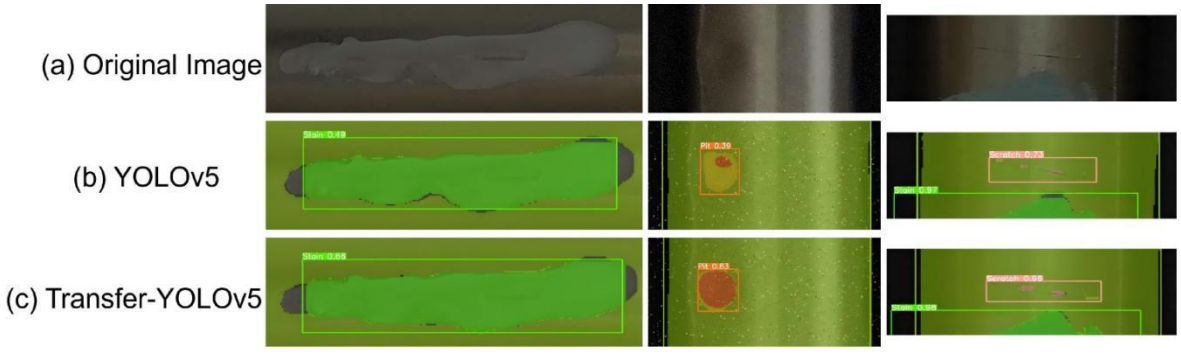

Figure. 5 YOLOv5s before and after optimization test result graphs. **(a)** Original defect images; **(b)** Detection results of YOLOv5s before optimization; **(c)** Detection results of Transfer-YOLOv5s after optimization. From left to right, the defect categories are stain, pit, and scratch. Produced by Python (version 3.7).

The significant improvement of the transfer learning-based YOLOv5s network in detecting scratch defects on batteries can be attributed to its ability to address challenges such as the subtle nature of scratches, low contrast, and appearance deformation through cross-domain feature alignment and hierarchical knowledge transfer. As the most common type of defect on battery appearances, scratches exhibit visual characteristics such as an elongated morphology, blurred edges, and susceptibility to interference from appearance reflections. These properties make it difficult for a conventionally trained YOLOv5s model to capture their discriminative features — this is because randomly initialized convolutional kernels in the early training stages tend to learn more prominent textures and contour features, while the ability to perceive the subtle gradient changes required for scratches relies on gradual optimization with a large number of high-quality samples. By introducing pre-trained weights from a source domain, transfer learning enables the model's initial parameters to encode generic low-level feature extraction capabilities for metal appearance defects. These fundamental visual units exhibit spatial correlations with the linear structural features of battery scratches. For instance, crack or scratch-like defects in the source domain share similar patterns in local gradient distributions with the target domain scratches, thereby significantly accelerating the model's adaptive fine-tuning to scratch features.

3.2. High-standard Detection Model Developed and Adaptability

In the **Figure. 6**, although the model converges under the current training configuration, its detection performance still shows notable shortcomings for tiny and low-visibility defects. The evaluation results indicate that the traditional NMS algorithm does not preserve enough information about subtle appearance variations, which provides the basis for further model optimization in the following section.

After optimization, as shown in the **Figure. 6 (c)**, Mask R-CNN exhibits consistent improvements across multiple metrics. The mAP_{bbox} increases from 89.23% to 94.80%, the mAP_{mask} rises from 81.65% to 83.07%, the recall improves from 84.50% to 87.78%, and the precision increases from 96.32% to 97.44%. Meanwhile, the inference time increases only slightly from 0.109 s to 0.117 s per image, indicating that the accuracy gains are achieved with minimal additional computational overhead. These improvements are mainly attributed to replacing the traditional NMS algorithm with SoftNMS, which decays the confidence scores of highly overlapped proposals instead of directly removing them.

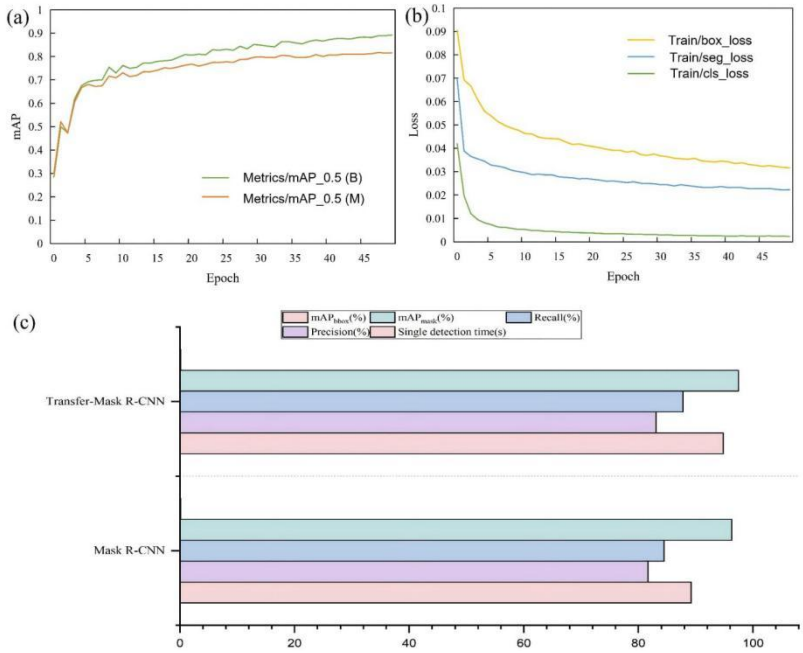

Figure. 6. Loss function and mAP curves during training. **(a)** mAP curve; **(b)** Loss function curve; **(c)** Analysis of the detection results of the model before and after optimization. Produced by Python (version 3.7) and Powerpoint (version 12.1.0.25865).

More importantly, the effect of SoftNMS is not uniform across all defect categories. The most evident improvement is observed for scratch defects and tiny low-visibility defect regions. These defects often generate multiple adjacent proposals with similar IoU and relatively low confidence, and the traditional NMS strategy tends to suppress true-positive candidates prematurely during post-processing. By preserving more plausible overlapping proposals, SoftNMS provides richer candidate information for the subsequent region of interest (RoI) based classification, bounding-box regression, and mask prediction stages. As a result, fine and continuous scratch regions can be retained more completely, and tiny subtle defects are more likely to be correctly recognized. In contrast, defect categories with clearer morphology and stronger contrast against the background, such as leakage or breakage, usually produce more stable candidate regions before optimization and therefore show relatively smaller gains. This indicates that the main contribution of SoftNMS lies in improving the detection of subtle, dense, and partially overlapping defects rather than uniformly enhancing all categories.

This category-level difference can be further explained by the characteristics of the optimized two-stage framework. Mask R-CNN first generates multi-scale candidate regions through the Region Proposal Network and then refines them via RoIAlign and the mask branch, which is advantageous for preserving subtle target information and delineating fine defect boundaries. For curved battery

appearances, a single scratch may appear as fragmented multi-segment projections in the image, and the traditional NMS algorithm tends to treat these overlapping boxes as redundant and suppress them. After replacing NMS with SoftNMS, more informative candidate regions can be retained, allowing the network to better preserve continuous scratch evidence and tiny defect details during post-processing. In addition, the backbone network combined with the feature pyramid structure integrates high-resolution low-level features with high-level semantic features, which is beneficial for capturing local gradient variations and weak boundary cues of tiny defects. Therefore, the adaptability enhancement of the optimized Mask R-CNN model is mainly reflected in its improved ability to parse complex, overlapping, and low-visibility defect patterns in high-precision inspection scenarios.

As shown in the **Figure. 7 (a-b)**, the model before optimization fails to capture several fine and overlapping scratch regions, often producing fragmented detection or low-confidence predictions. After optimization, more continuous scratch regions are detected with higher confidence, indicating that SoftNMS preserves informative overlapping candidates and allows the network to retain subtle defect evidence during post-processing. Similarly, **Figure. 7 (c-d)** show that the optimized model identifies tinier and finer defect regions that are missed by the baseline model. These qualitative results are consistent with the quantitative analysis and further confirm that the advantage of the optimized Mask R-CNN model is particularly prominent for scratch defects and tiny low-visibility defect subsets.

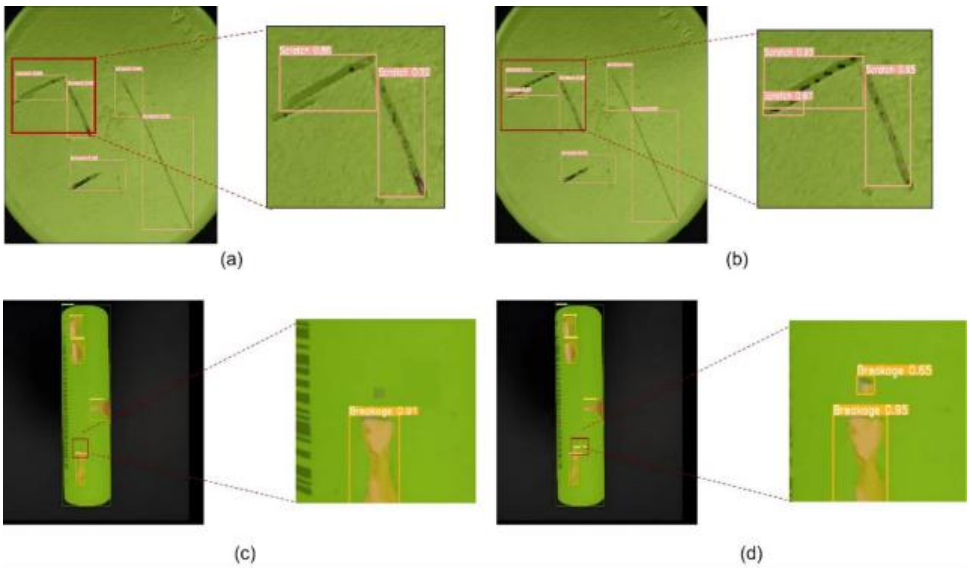

Figure. 7 Comparison of tiny target defect detection effect before and after model optimization. **(a) (c)** Mask R-CNN; **(b) (d)** Optimization-Mask R-CNN. Produced by Python (version 3.7).

3.3. Visualization Model Development

To make the detection results more intuitively displayed, a dedicated defect detection software system was designed and developed. Based on the system framework and detection workflow, three interfaces were designed: the user registration interface, the user login interface, and the defect detection system interface.

The user registration enabled the system to manage user information. It also managed accounts by recording registration details and supporting user categorization. By verifying the user's registration information, the system ensured the legitimacy of user identities. The user login screen was used to authenticate the user and restrict system access to authorized individuals. The interface of the defect detection system for echelon utilization consists of two main parts, which are five function keys and a detection effect window. The system interface is shown in the **Figure. 8**.

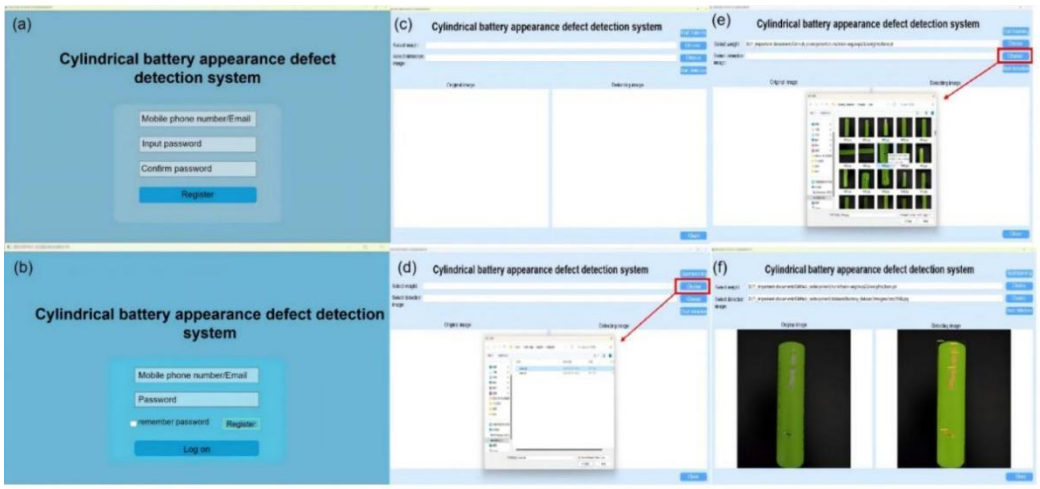

Figure. 8. System of the defect detection. **(a)** Registration interface; **(b)** Login interface; **(c)** Defect detection interface; **(d)** Choose weight paths; **(e)** Choose the image's path to be examined; **(f)** Test results interface. Produced by Python (version 3.7) and Powerpoint (version 12.1.0.25865).

The system was able to detect appearance defects and present the results clearly on the interface. Users could view the defect category, its position on the battery appearance, and the corresponding confidence score. In addition, the system allowed basic operations such as image importing, result saving, and repeated testing, which made the system practical for real use. These functions demonstrate that the proposed framework is not limited to algorithmic verification but can serve as a practical component in early-stage battery detection workflows. The system developed in this work exhibits promising applicability and practical value in real production settings.

4. Conclusion

This work creates a representative dataset of cylindrical lithium-ion battery appearance defects for pre-echelon defect detection. The YOLOv5 model was optimized with transfer learning, improving detection performance and efficiency, and achieving 96.70% precision, 84.30% recall, and 93.40% mAP, with results delivered in 5.5 ms per image. However, tiny and overlapping defects remained challenging. To address this issue, a two-stage Mask R-CNN model was introduced for high-precision defect inspection, in which the traditional NMS algorithm was replaced with the SoftNMS algorithm, boosting precision to 97.44% and reducing missed detection. In contrast to YOLOv5, which is suitable for fast online screening because of its real-time detection capability, Mask R-CNN is more appropriate for inspection scenarios with stringent accuracy requirements. A PyQt5-based real-time monitoring system was further developed for intuitive human-machine interaction and industrial deployment. The system integrates dataset creation, model optimization, and real-time visualization, providing a practical solution for intelligent defect detection.

Nevertheless, this work still has several aspects that can be further improved. First, although the dataset was expanded to 3,750 images through augmentation, the number of retained original images is still limited, and the current dataset may not fully cover the inter-batch variability of retired batteries. Future work should therefore introduce cross-batch validation using battery samples collected from different recycling batches, suppliers, and operating histories, so as to better evaluate model generalization under realistic source variation. Second, the present work is based on static image inspection, whereas actual industrial sorting is often performed in continuous conveyor or video-stream scenarios. Future research should extend the current framework to video-based inspection, with emphasis on temporal consistency, frame-to-frame stability, and real-time defect tracking. Third, the current experiments were mainly conducted under relatively controlled imaging conditions. To improve industrial adaptability, future studies should test the models under more complex conditions, such as variable illumination, reflective interference, motion blur, occlusion, background clutter, and mixed defect distributions. Fourth, the evaluation in this work mainly focused on precision, recall, and mAP. Future work will extend the proposed framework from pre-echelon defect inspection to broader industrial cell screening, such as the categorization of production-side A- cells, B- cells, and other low-quality cells. We also plan to introduce fine-grained indicators like F1-score, IoU-related statistics, category-level robustness analysis, and error distribution analysis for tiny and overlapping defects. Additionally, incorporating multi-sensor information such as depth, infrared, or multi-view imaging could enhance the identification of subtle defects that are difficult to distinguish using visible-light imaging alone. To further validate the proposed approach, horizontal comparisons with other commonly used detection models, such as YOLOv4, Faster R-CNN, and RetinaNet, will be included, assessing the performance across different models and improving its applicability in real-world scenarios. Furthermore, integrating computed tomography (CT) imaging will provide an additional layer of defect detection, enabling the identification of internal defects in batteries that appear visually intact.

Abbreviation List

YOLO: You Only Look Once
Mask R-CNN: Mask Region-based Convolutional Neural Network
NMS: Non-Maximum Suppression
SoftNMS: Soft Non-Maximum Suppression
IoU: Intersection over Union
AP: Average Precision
mAP: mean Average Precision
FPS: Frames Per Second
mAP_{bbox}: mean Average Precision for bounding boxes
mAP_{mask}: mean Average Precision for masks

Author Contribution

We confirm that all listed authors meet all four authorship criteria. Specifically, Yuhui Peng, Yaxin Chi, Xihua Zhang, Tao Zhang, Xuning Zhuang, Xiaolong Song have made substantive contributions to the conception and design of the study, data acquisition, analysis, or interpretation; participated in drafting the manuscript or revising it critically for important intellectual content; approved the final version of the manuscript; and agreed to be accountable for all aspects of the work to ensure that any questions related to the accuracy or integrity of the work are appropriately investigated and resolved.

Supplementary Information

Supplementary material associated with this article can be found, in the online version, at **Supplementary Information**.

Conflicts of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was financially supported by the National Natural Science Foundation of China (No. 42471331) and Shanghai Pudong New Area Livelihood Research Project (No. PKJ2023-C03).

Artificial Intelligence (AI) Statements

The authors acknowledge the use of artificial intelligence tools, such as ChatGPT, for language editing and refinement of certain sections of the manuscript. The AI tools were utilized to assist with sentence structure, grammar correction, and overall readability, ensuring a clear and coherent presentation of the research findings. The authors affirm that the AI tools were used to assist with language-related aspects only, and all content, analyses, and conclusions are

solely the authors' original work.

References

[1] Liu C, Huang S, Fu Z, *et al.* Recycling and Echelon Utilization of Used Lithium-Ion Batteries from Electric Vehicles in China. International Journal of Electrochemical Science. 2022, 17 (6). [Google Scholar] [CrossRef]

[2] Zeng J. Technical-economic analysis for cascade utilization of spent power batteries in the energy storage system. Journal of Energy Storage. 2025, 114. [Google Scholar] [CrossRef]

[3] Tian GD, Yuan G, Aleksandrov A, *et al.* Recycling of spent Lithium-ion Batteries: A comprehensive review for identification of main challenges and future research trends. Sustainable Energy Technologies and Assessments. 2022, 53. [Google Scholar] [CrossRef]

[4] Ma R, Chen J, Feng Y, *et al.* ELA-YOLO: An efficient method with linear attention for steel surface defect detection during manufacturing. Advanced Engineering Informatics. 2025, 65. [Google Scholar] [CrossRef]

[5] Liu L, Du D, Sun Y, *et al.* SFMW-YOLO: A lightweight metal casting surface defect detection method based on modified YOLOv8s. Expert Systems with Applications. 2025, 287. [Google Scholar] [CrossRef]

[6] Sun P, Hua C, Ding W, *et al.* A real-time detection framework for surface defects in ceramic tableware based on deep learning. Expert Systems with Applications. 2025, 286. [Google Scholar] [CrossRef]

[7] Nie J, Wang Y, Yu Z, *et al.* High-precision grain size analysis of laser-sintered Al2O3 ceramics using a deep-learning-based ceramic grains detection neural network. Computational Materials Science. 2025, 250. [Google Scholar] [CrossRef]

[8] Zhang WA, Tian CH, Yang JF, *et al.* Vision and deep learning-based tunnel surface defect detection: Technological evolution, challenges, and prospects. Measurement. 2026, 257. [Google Scholar] [CrossRef]

[9] Chen Z, Chen T, Zheng H, *et al.* Lightweight deep learning based solar cells defect detection using electroluminescence images. Engineering Applications of Artificial Intelligence. 2025, 162. [Google Scholar] [CrossRef]

[10] Wang J, Sun A, Wang C, *et al.* Improved deep learning-based detection of drainage pipeline defects using an enhanced YOLOv8 framework. Journal of Water Process Engineering. 2025, 78. [Google Scholar] [CrossRef]

[11] Wang ZW, Hu ZC, Yang WF, *et al.* A chip-wound inductors defect detection method based on object detection and semantic segmentation. Measurement. 2025, 250. [Google Scholar] [CrossRef]

[12] Li YK, Zhang XM. Lightweight deep learning model for underwater waste segmentation based on sonar images. Waste Management. 2024, 190: 63-73. [Google Scholar] [CrossRef]

[13] Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. 2016: 779-788. [Google Scholar] [CrossRef]

[14] Wei L, Dragomir A, Dumitru E, *et al.* SSD: Single Shot MultiBox Detector. Springer, Cham. 2016: 21-37. [Google Scholar] [CrossRef]

[15] Jiang WJ, Li TF, Zhang SL, *et al.* PCB defects target detection combining multi-scale and attention mechanism. Engineering Applications of Artificial Intelligence. 2023, 123. [Google Scholar] [CrossRef]

[16] Zuo DD, Ge JQ, Yu YM. Surface Defect Detection of Button Cells Based on the Fusion of Object Detection and Semantic Segmentation. Arabian Journal for Science and Engineering. 2025, 50 (22). [Google Scholar] [CrossRef]

[17] Girshick R. Fast R-CNN. IEEE International Conference on Computer Vision (ICCV). 2015: 1440-1448. [Google Scholar] [CrossRef]

[18] Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2014: 580-587. [Google Scholar] [CrossRef]

[19] He K, Gkioxari G, Dollár P, *et al.* Mask R-CNN. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020,42 (2): 386-397. [Google Scholar] [CrossRef]

[20] Ren S, He K, Girshick R, *et al.* Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017,39 (6): 1137-1149. [Google Scholar] [CrossRef]

[21] Xie Q, Zhou W, Tan H, *et al.* Surface Defect Recognition in Steel Plates Based on Impoved Faster R-CNN. 2022 41st Chinese Control Conference (CCC). 2022: 6759-6764. [Google Scholar] [CrossRef]

[22] Cha Y J, Choi W, Suh G, *et al.* Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detecting Multiple Damage Types. Computer-Aided Civil and Infrastructure Engineering. 2018, 33 (9): 731-747. [Google Scholar] [CrossRef]

[23] Janiesch C, Zschech P, Heinrich K. Machine learning and deep learning. Electronic Markets. 2021, 31 (3): 685-695. [Google Scholar] [CrossRef]

[24] Lu L, Sun N, Wang Z, *et al.* Tile detection using mask R-CNN in non-structural environment for robotic tiling. Automation in Construction. 2025,171. [Google Scholar] [CrossRef]

[25] Torralba A, Russell BC, Yuen J. LabelMe: Online Image Annotation and Applications. Proceedings of the IEEE. 2010,98 (8): 1467-1484. [Google Scholar] [CrossRef]

[26] Zhuang FZ, Qi ZY, Duan KY, *et al.* A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE. 2021,109 (1): 43-76. [Google Scholar] [CrossRef]

[27] Bao Y, Song K, Liu J, *et al.* Triplet-Graph Reasoning Network for Few-Shot Metal Generic Surface Defect Segmentation. IEEE Transactions on Instrumentation and Measurement. 2021,70: 1-11. [Google Scholar] [CrossRef]

[28] Song KC, Yan YH. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. Applied Surface Science. 2013,285: 858-864. [Google Scholar] [CrossRef]